# Automatic Spectrum Recognition System for Charge State Analysis in Electron Cyclotron Resonance Ion Sources

Rui Wang [1,2] Cheng Qian [1] Yu-Hui Guo [1,*] Peng Zhang [1] Jin-Dou Ma [1]

**Affiliations:**

[1]Institute of Modern Physics, Chinese Academy of Sciences, Lanzhou 730000, China

[2]University of Chinese Academy of Sciences, Beijing 100049, China

[*]Corresponding author. *E-mail address:* guoyuhui@impcas.ac.cn

**Abstract:** The Electron Cyclotron Resonance (ECR) ion source is a critical device for producing highly charged ion beams in various applications. Analyzing the charge-state distribution of the ion beams is essential, but the manual analysis is labor-intensive and prone to inaccuracies due to impurity ions. An automatic spectrum recognition system based on intelligent algorithms was proposed for rapid and accurate chargestate analysis of ECR ion sources. The system employs an adaptive window-length Savitzky-Golay (SG) filtering algorithm, an improved automatic multiscale peak detection (AMPD) algorithm, and a greedy matching algorithm based on the relative distance to accurately match different peaks in the spectra with the corresponding charge-state ion species. Additionally, a user-friendly operator interface was developed for ease of use. Extensive testing on the online ECR ion source platform demonstrates that the system achieves high accuracy, with an average root mean square error of less than 0.1 A for identifying charge-state spectra of ECR ion sources. Moreover, the system minimizes the standard deviation of the first-order derivative of the smoothed signal to 81.1846 A. These results indicate the capability of the designed system to identify ion beam spectra with mass numbers less than Xe, including Xe itself. The proposed automatic spectrum recognition system represents a significant advancement in ECR ion source analysis, offering a rapid and accurate approach for charge-state analysis while enhancing supply efficiency. The exceptional performance and successful implementation of the proposed system on multiple ECR ion source platforms at IMPCAS highlight its potential for widespread adoption in ECR ion source research and applications.

**Key words:** ECRIS Spectrum recognition SG filtering AMPD algorithm Greedy algorithm



## 1 Introduction

The ECR (Electron Cyclotron Resonance) ion source, initially developed by Prof. Geller and colleagues at Grenoble Laboratory in France [1, 2], has evolved into a highly efficient and highly charged state ion source with a diverse range of beam types. It is renowned for its exceptional stability and reproducibility and has found widespread application in heavy ion acceleration facilities worldwide. For instance, the VENUS ion source at Lawrence Berkeley National Laboratory (LBNL) provides beam current for the 88-inch cyclotron [3], the SC-ECR ion source at the Institute of Physical and Chemical Research (RIKEN) in Japan supplies beam current for the Radioisotope Facility (RIBF) [4], and the SECRAL ion source at the Institute of Modern Physics (IMP) in Lanzhou, China, serves as the beam current provider for the Heavy Ion Research Facility (HIRFL) [5]. The beam transmission lines of the ECR ion source are illustrated in Fig. 1, where ions of varying charge-to-mass ratios ( $q/M$ ) exhibit distinct deflection trajectories in the dipole magnet, resulting in different peak positions in the spectrum. This property allows for separating and sorting ions based on their charge-to-mass ratio, a crucial aspect of the transport system in analyzing and manipulating ion beams in ECR ion source research and applications. The spectrogram of the beam reflects the energy distribution of the ions. Accurate and rapid identification of the charge-state of the injected ion beam is crucial

for optimizing the beam supply efficiency of the ion source. However, the traditional manual identification method, relying solely on the experience of engineers, entails complex calculations and comparisons and is susceptible to misidentification or omission due to system background noise. Furthermore, manual spectrum recognition has limitations such as a high technical threshold, prolonged time consumption, and a low accuracy rate [6-9]. To address these challenges, this paper presents a novel automatic spectrum recognition system for beam charge-state distribution spectra, featuring three key algorithms: the Savitzky-Golay (SG) filtering algorithm with adaptive window length, the improved automatic multiscale peak detection (AMPD) algorithm, and the greedy matching algorithm based on relative distance. These intelligent algorithms collectively enable fast and accurate spectra recognition while reducing human involvement, lowering the technical threshold, and offering significant research value.

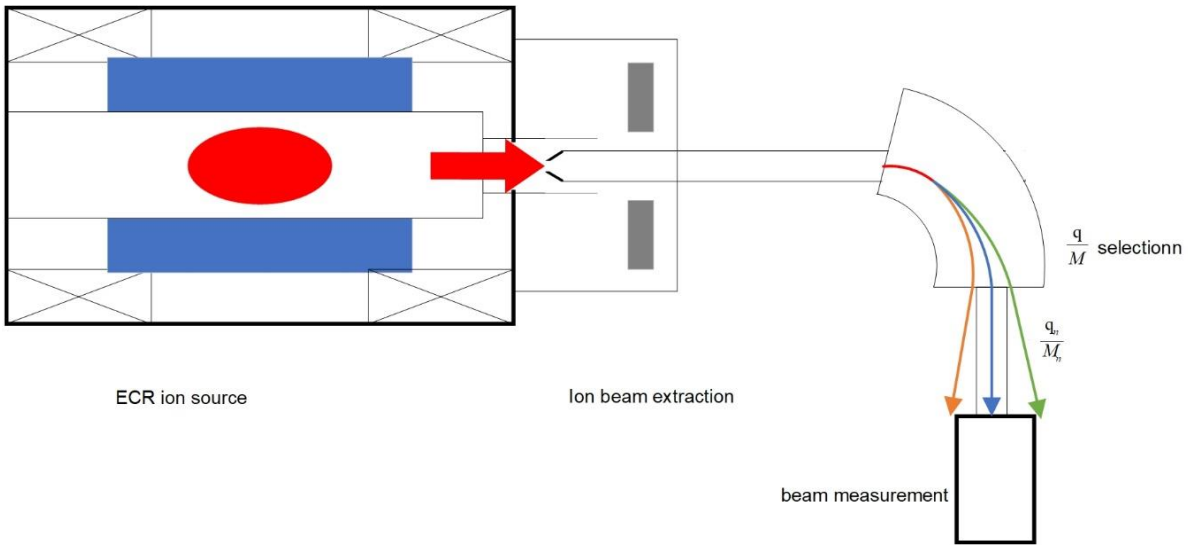


**Fig. 1** ECR Beam Transport System. In this system, the beam deflection was achieved using a dipole magnet. The trajectory of the deflected beam varied based on the charge-to-mass ratio of the ions. The dipole magnet induced a magnetic field that caused the charged particles in the beam to experience a Lorentz force, resulting in a curved path. The curvature of the trajectory depends on the charge-to-mass ratio of the ions, with lighter ions exhibiting larger curvatures than heavier ions.

Influenced by environmental background noise, the spectrogram signal of the beam is often contaminated with noise. Several filtering algorithms have been proposed to mitigate this issue, including wavelet denoising [10], empirical modal decomposition (EMD) [11], Savitzky-Golay (SG) filtering [12], Legendre filtering [13], kernel regression [14], local extremum center [15], and signal sparsity-based denoising [16]. Among these, SG filtering has been widely utilized in diverse fields, such as digital control systems [17], electrocardiogram denoising [18], and nuclear electrical reaction calculations [19]. Preserving the signal shape and peak properties in SG filtering [20] makes it particularly appealing for the current study based on previous research. SG filtering is known to effectively filter noise and outliers while preserving the underlying trend and periodicity of the signal.

The first crucial step in spectrogram identification is the accurate and comprehensive detection of spectral peaks. Peak detection in signal processing plays a pivotal role in obtaining reliable results. Various algorithms have been developed for peak detection, including traditional window-threshold methods [21], wavelet transforms [22], template techniques [23], hidden Markov models [24], and others. However, most of these peak-finding algorithms suffer from the challenge of setting hyperparameters before their application. In practice, it has been observed that many commonly used algorithms require a significant number of hyperparameters to be pre-set, leading to varying results in peak detection even within the same spectrum due to different combinations of window length and peak height threshold. Determining the optimal parameter combination is not universally applicable, making the tuning process laborious and challenging. An alternative approach that addresses these limitations is the automatic

multiscale peak detection (AMPD) algorithm, which sets multiple window scales for peak detection based on the input signal characteristics. This method does not rely on prior knowledge or require frequent tuning, making it more suitable than other algorithms for detecting peaks in a spectrum.

The identification of beam spectrograms can be achieved by matching the identified peaks with the beam charge states [25]. The performance of the matching algorithm directly affects the accuracy of spectrogram identification results. A commonly used approach is the greedy matching algorithm, which determines the optimal local resolution for each subproblem and combines the results to obtain a final global solution [26]. The essential advantage of this algorithm lies in its high efficiency and ability to quickly obtain a global solution through locally optimal solutions [27]. Greedy algorithms have been widely employed in various fields, such as task scheduling problems in optimization [28], minimum spanning tree problems in graph theory [29], and queuing problems in queuing theory [30]. Despite the effectiveness of greedy algorithms in obtaining locally optimal solutions, it is important to note that they may not always find globally optimal solutions due to their lack of backtracking. However, greedy algorithms generally yield satisfactory results and are commonly used in practical applications.

This study presents a novel approach for denoising spectrograms by introducing an adaptive window length SG filtering algorithm. The proposed algorithm can effectively remove background noise from the original spectrogram. By automatically determining the optimal window length, our algorithm overcomes the limitations of existing methods that require manual hyperparameter tuning, making it more practical and user-friendly. In addition, an improved AMPD algorithm is developed for spectral peak calibration. One notable advantage of our algorithm is its ability to automatically eliminate false peaks without needing prior hyperparameter settings, improving peak detection accuracy and reliability in spectrograms. Furthermore, the limitation of the greedy algorithm in global optimization is addressed by proposing an interquartile range (IQR) anomaly detection mechanism [31] based on relative distance. This mechanism aims to identify and reject solutions in the solution set that satisfy the greedy strategy but are outliers in the global context. By mitigating the limitations of the greedy algorithm, the accuracy and robustness of our algorithm in identifying spectral peaks are enhanced. The accuracy of the final identification results is evaluated and verified in this paper using a discriminative empirical formula. The experiment results demonstrate the effectiveness of our proposed algorithms in achieving high-quality denoising, accurate peak detection, and precise charge identification in spectrograms. The automatic spectrum recognition system presents a promising approach to analyzing spectrograms in various applications.

## 2 Architecture of Automatic Spectrum Recognition System

A human-computer interface is built into this system, and the operator can specify and send down the relevant parameters. After receiving the parameters from the front end, the back-end program will first scan the spectrum according to the start current, the stop current, and the scan step. Two critical parameters must be collected in the scanning process: the dipole magnet current and the Faraday cup intensity. In this paper, the dipole magnet current is acquired through a resistive shunt using an AMETEK SGA high-power programmable DC power supply [32] with a current acquisition range of 5 A to 6000 A. The Faraday cup intensity is acquired using an Agilent 34410A high-performance digital multimeter with a "6 1/2" resolution and a sampling frequency of 1 KHz [33]. The dipole magnet current values are rounded to four decimal places in amperes (A), and the Faraday cup intensity values are rounded to two decimal places in electrostatic microamperes (eμA) to ensure ease of calculation and legibility. The collected dipole magnet currents and Faraday cup intensities are transmitted to the data processing module. The signal is first smoothed and filtered in this module using the Savitzky-Golay (SG) filtering algorithm with adaptive windows, after which the smoothed signal is searched for using an improved automatic multiscale peak detection (AMPD) algorithm. Moreover, a greedy matching algorithm based on relative distance is proposed to achieve the matching between peak and ion charge states. Finally, the final matching results are visually outputted after outlier rejection. The implementation of each algorithm will be described in detail in the following. The overall system architecture is shown in Fig. 2. The final results are displayed on the Qt interface, a widely used graphical user interface (GUI)

toolkit for creating interactive software applications [34]. The data is processed and displayed by the Qt interface clearly and understandably to enable the system's effective operation, allowing the operator to monitor the data processing results concisely and change the parameters in time for sending.

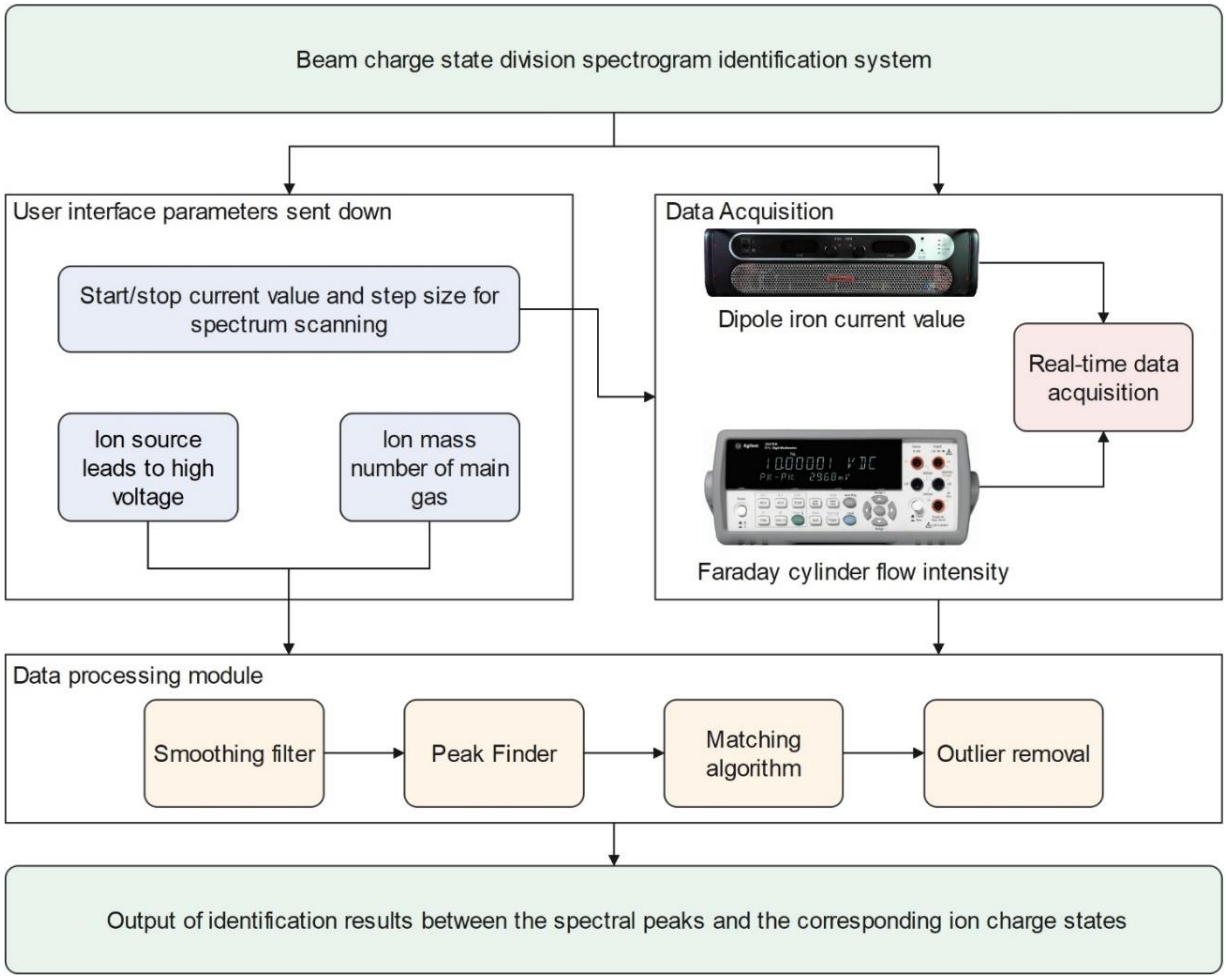


**Fig. 2** The overall architecture of the beam charge-state distribution spectrogram identification system. Once the parameters specified by the user through the human-computer interface are acquired, the data is obtained in real time by reading and writing the process variable (PV) value. The acquired data, including the dipole magnet currents and Faraday cup intensities, is then processed using the data processing module.

The beam intensity is measured using a high-precision multimeter, while the dipole magnet power supply is responsible for collecting the current of the dipole magnet with an original hardware sampling rate of 1 KHz. Data downsampling is necessary to ensure synchronized data and to accommodate the architecture of the ECR ion source control system. After issuing a dipole current write command during a set sweep step, the dipole magnet current and Faraday cup intensity values were read back after a delay of 0.1 seconds. This interval allowed the dipole magnet sufficient time to respond to the control command. Consequently, the sampling rate is reduced to 10 Hz to ensure that the dipole magnet current and Faraday cup intensity match [35]. This process is demonstrated in Fig. 3.

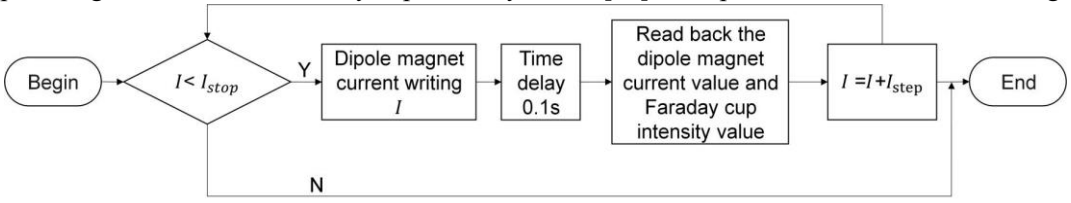


**Fig. 3** Data acquisition and synchronization process.

### 2.1 Design of SG filtering algorithm with adaptive window length

According to the design characteristics of the SG filtering algorithm, a larger window length is associated with improved robustness against noise rejection and reduced variance in the error deviation for a given filter order. However, excessively large window lengths can result in distortion and bias in the filter output compared to the actual signal. Therefore, finding the optimal parameters for the SG filtering algorithm involves balancing bias and variance in the error estimation. The parameters of the SG algorithm with an adaptive window length include the order $n$ of the polynomial and window length of $N$=2M+1. Within each window, a cost function at the following should be minimized, given by Eq. (1).

$$\delta_{\mathrm{n}} = \sum_{i=-M}^{M} (\sum_{\mathrm{k}=0}^{n} a_k i^k - x_i)^2 \tag{1}$$

In Eq. (1), $a_k$ represents the $k$ th polynomial coefficient. In this case, the output of SG filtering can be expressed as:

$$y(k) = \sum_{j=-\frac{N-1}{2}}^{\frac{N-1}{2}} W_n(j) x_{k-j} \tag{2}$$

where $W_n(j)$ is the $j$ th sample of the continuous function $W_n(z)$ and $W_n(z)$ is a polynomial that can be defined as

$$W_n(z) = a_{n+1} \frac{q_{n+1}(z)}{z} \tag{3}$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{4}$$

According to Eq. (4), $\binom{n}{k}$ represents the binomial coefficient. The equivalence of $a_{n+1}$ in this context is expressed by Eq. (5):

$$a_{n+1} = \frac{n+1}{2^{n+1}} \binom{n}{\frac{n}{2}} \frac{\sqrt{(-1)^n}}{N\left(N^2-2^2\right)\cdots\left(N^2-n^2\right)} \tag{5}$$

Similarly, $q_n(z)$ is a shifted Chebyshev polynomial obtained from the $n$ th forward difference $\Delta^n$. This instance is shown in Eq. (6).

$$q_n(z) = n!\Delta^n\left[\binom{z+M}{n}\binom{z-M-1}{n}\right] \tag{6}$$

An SG filter of order $n$ and window length $N$ is used to reconstruct the signal $f(t)$ contaminated with noise. When the signal $f(t)$ is sufficiently smooth such that there are $n+2$ consecutive derivatives, the optimal window length of the SG filter [36] can be approximated using Eq. (7) as follows:

$$N_{opt} = \sqrt[2n+5]{\frac{2(n+2)\left((2n+3)!\right)^2}{\left((n+1)!\right)^2}\frac{\sigma^2}{v_n}} \tag{7}$$

where $\sigma^2$ is the variance, and $v_n$ can be considered a function of the signal correlation, where $v_n/\sigma^2$ can be interpreted approximately as the signal-to-noise ratio (SNR); the lower SNR to a larger optimal window length.

In this algorithm, the parameters utilized in this process encompass the original signal input denoted as $x = [x_1, x_2 \ldots, x_L]^T$, the window length $N_{opt}$, the filter order $n$, as well as defining the filter output as

$y=\left[y_1, y_2 \ldots, y_L\right]^T$. To calculate $v_n$, the original signal $x$ is filtered by SG as the initial step. Because of the reduced computational workload, the original window length is defined to be five here, and the filter order is two. The resulting output $y$ after the first SG filter is obtained as its first-order derivative. Subsequently, SG smoothing is iteratively performed, with the smoothed window length denoted as $N_1=2\left\lfloor N_{opt}/2\right\rfloor$. The iterative application of the SG filtering algorithm is crucial in enhancing the estimation accuracy of the desired parameter $v_n$. By repeatedly filtering the original signal with SG, the algorithm refines the estimate of $v_n$, leading to a more precise and reliable result. The optimal window length $N_{opt}$ determined through an iterative process based on Eq. (7). The iterative evaluation of $N_{opt}$ continues until $N_1=N_{opt}$ is achieved, indicating the termination of the process. This adaptive approach ensures that the SG filtering algorithm adjusts the window length based on the specific characteristics of the input signal, resulting in an optimized window length that enhances the accuracy of the filtered signal.

### 2.2 Design of the improved AMPD algorithm

The improved AMPD algorithm proposed in this study enhances the original AMPD algorithm by incorporating constraints on the peak output. The original AMPD algorithm employs a moving window method to find the local maxima of the signal, where the window length, denoted as $w_k$, can be expressed as $\left\{w_k=2k \mid k=1,2\cdots, L\right\}$, where $L$ is a parameter associated with the range of the signal [37]. However, if $L$ is set excessively large, certain spectral peaks may remain unidentified. On the other hand, choosing $L$ as too small may result in the inclusion of "false peaks" in the results.

To ensure the complete detection of all spectral peaks of interest, the parameter $L$ is designed in this study as $L=\lceil N/10\rceil-1$, where $N$ represents the signal range. While the results accurately identify all spectral peaks of interest, they may also contain spurious peaks. For instance, during beam pauses or when there is a lack of beam flow, the Faraday cup intensity in the optimal state should ideally be 0 eμA. However, due to the influence of disruptions from the device itself, the monitored Faraday cup intensity at this time may fluctuate slightly above and below 0 eμA. Nonetheless, these false peaks are less significant compared to the spectral peaks formed by the beam current, which requires attention. As the peak positions and peaks in the spectrograms of different types of beams may vary considerably, it was not feasible to establish a general significance level threshold. To avoid introducing additional hyperparameters into the algorithm, the original AMPD algorithm's peak output is improved in this study by specifying that only peaks with a significance level higher than 2% of the spectrogram based on the highest peak value are considered beam spectrum peaks. This empirical threshold helps filter out false peaks and retain only the peaks most likely to represent actual spectral peaks of interest. The process of the algorithm at this stage is as follows: firstly, the local maximum scale (LMS) is calculated based on parameter $L$. Secondly, the row-wise sum of the LMS is computed, and the LMS is reconstructed based on the minimum value of the row sum. In the third step, the peak-seeking output of the AMPD algorithm is obtained. Finally, a new peak sequence is re-output after filtering the output peaks according to the empirical threshold. The calculation process is illustrated in Fig. 4:

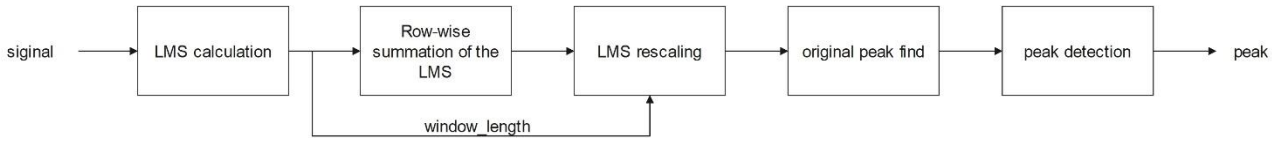


**Fig. 4** Calculation steps of the improved AMPD algorithm

### 2.3 Design of the Greedy algorithm based on the relative distance

In this study, a greedy matching algorithm based on the relative distance was developed to address the issue of matching peaks to charge states. This algorithm follows a greedy strategy compatible with the physical properties of the ion source. The difference between the peak detected in the previous section and the calculated theoretical value (according to Eq. (8)) was determined. The matches that met the requirements for each peak were identified by

screening according to the greedy strategy, and greedy matching was completed based on this screening process. Considering the limitations of the greedy algorithm, which only involves local optimal solutions and cannot be backtracked, the concept of relative distance is introduced in this study. As the change in the charge-state of each ion in the same spectrogram is usually continuous, the first recognition result of the greedy algorithm was classified according to the ion species, and the coded charge states of the classified ions were obtained. Finally, the first recognition result is further refined using interquartile range (IQR) anomaly detection to reject outliers and obtain the final recognition result. This step helps improve the accuracy and reliability of the charge-state identification process. The specific process of the algorithm used in this study is as follows:

The standard peak $C$ of the total charge-state was initially calculated based on the empirical formula of the ion source within a specified sweep range, using Eq. (8). In this equation, $\mu$ represents the constant of the dipole magnet resolution system of the ECR ion source, $U$ denotes the extracted high voltage, and $M/q$ represents the ion mass-to-charge ratio.

$$C = \mu\sqrt{U * \frac{M}{q}} \tag{8}$$

The set of peaks computed in the previous section is denoted as $P = \{P_1, P_2 \cdots P_p\}$, and the set of charge-state standard peaks is denoted by $C = \{C_1, C_2 \cdots C_c\}$. These two sets are encoded separately. The next step is to devise a greedy strategy that generates a set of solutions that adhere to the greedy strategy. In this paper, the chosen greedy strategy is based on the resolution of the dipole magnet against the spectral peaks, and it is defined as follows, where $s$ means the minimum distance index:

$$\left|\frac{M}{q} - \frac{M_s}{q_s}\right| < 0.1 \tag{9}$$

Once an appropriate greedy strategy is determined, the next step is to calculate the relative distance between each peak and its corresponding theoretical value. The results were organized into a relative distance matrix, denoted as $L$ (Eq. (10)), where $L_{(p,c)} = \left|P_p - C_c\right|$.

$$L = \begin{bmatrix} L_{(1,1)} & \cdots & L_{(1,c)} \\ \vdots & \ddots & \vdots \\ L_{(p,1)} & \cdots & L_{(p,c)} \end{bmatrix} \tag{10}$$

The $i$-th peak, which corresponds to the minimum distance index $s$, is filtered based on $L$, where $L_{(i,s)} = argmin\left(L_{(i,1)}, L_{(i,2)} \cdots L_{(i,c)}\right)$. Then, the set of solutions $R_{\text{old}} = \{R_1, R_2 \cdots R_p\}$ is expanded according to the greedy strategy, where $R_i = \{L_{(i,s)} \cdots L_{(i,n)}\}$.

After obtaining the set of raw solutions, the next step in the algorithm is to remove outliers from this set. First, the original codes were classified according to the ion species and then normalized using the robust_stable method. This method eliminates the magnitude of the data and makes them comparable [38]. This helps to account for any variations in magnitude or scale among the different ion species, allowing for more robust outlier detection. Finally, the outliers were rejected after anomaly detection using the IQR, as described in Eq. (11). The IQR is a measure of the dispersion or spread of data and is calculated as the difference between the 75th ($Q_3$) and 25th percentiles ($Q_1$) of the data [39]. This outlier rejection step helps remove any spurious or erroneous solutions from the data, ensuring that only reliable and accurate solutions are considered in the subsequent analysis or processing steps of the algorithm.

$$\begin{cases} IQR = Q_3 - Q_1 \\ Q_{min} = Q_1 - \mathrm{k}IQR \\ Q_{max} = Q_3 + \mathrm{k}IQR \end{cases} \tag{11}$$

In Eq. (11), $Q_{\max}$ represents the upper quartile and $Q_{\min}$ represents the lower quartile. When an observation does not satisfy $Q_{min} < Q < Q_{max}$, it is considered an outlier and must subsequently be removed from the solution set [40]. The final solution set is $R_{\text{new}} = \left\{ R_{\text{new1}}, R_{\text{new2}} \cdots R_{newp} \right\}$. Fig. 5 illustrates the flow schematic of the greedy matching algorithm based on relative distance. The detailed procedure for this algorithm is presented in Algorithm 1. This algorithm utilizes the concept of relative distance and incorporates the IQR method for outlier rejection to improve the accuracy and reliability of the matching process for identifying the charge states [41] of ions in the spectrogram.

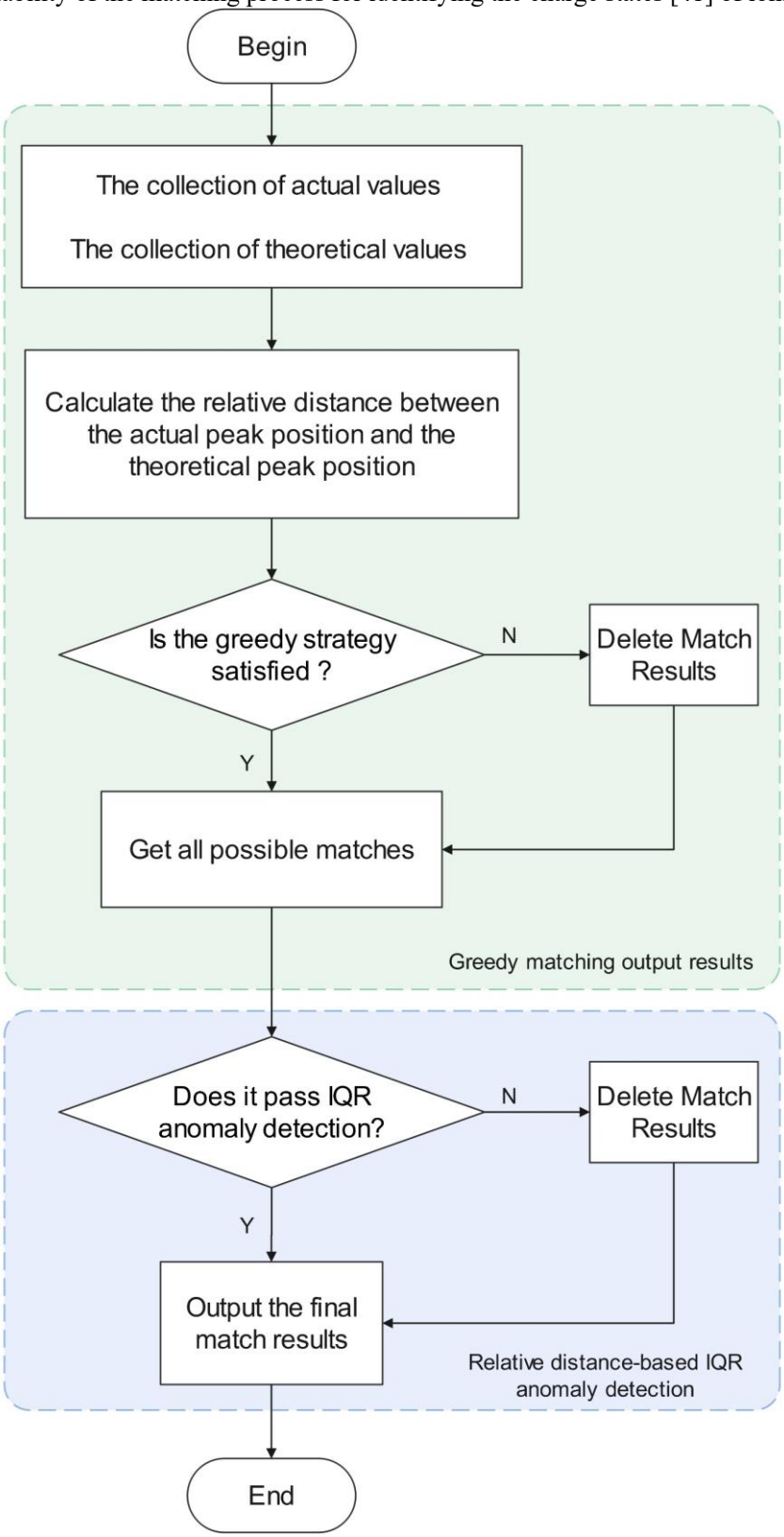

**Fig. 5** The flow diagram of the greedy matching algorithm based on relative distance

**Algorithm 1** Greedy algorithm based on relative distance

**Input:** Collection of computed values A, Collection of theoretical values B, All encoding results OLDInd, Threshold λ

**Output:** Corrected match result NEWInd

When the difference between the mass-to-charge ratios of the two particles is less than a threshold λ, it is considered that multiple elements in the spectrum belong to the same peak. They belong to the same match.
i=getColumNum(A)
j=getColumNum(B)
**for** A[i] in A **do**
  **for** B[j] in B **do**
    abs(A[i] - B[j])= d[i][j]
    Distance = d[i][j]
  **end for**
**end for**
**if** Distance <λ **then**
  num[i]=j
  Ind=num
  Index=OLDInd[i][Ind]
**else**
  continue;
**end if**
**while** isNotEmpty(Index[i]) **do**
  $n = len(\text{Index})$
  $Q_1 = \frac{n+1}{4}; Q_3 = \frac{3(n+1)}{4}$
  $IQR = Q_3 - Q_1$
  initial k ▷ k is the set anomaly index
  **if** $Q_1 - kIQR < \text{Index}[i][j] < Q_3 + kIQR$ **then** ▷ Remove outliers
    NEWInd[i] ←Index[i][j]
  **end if**
**end while**
**return** NEWInd

## 3 Automatic Spectrum Recognition System Performance Evaluation and Operational Testing

### 3.1 Performance analysis of SG filtering algorithm with adaptive window length

The present study uses accurate Ar and Xe beam data to evaluate the proposed algorithm. The results of applying the SG filter algorithm with an adaptive window length to the charge-state distribution spectra of the Ar and Xe beams are presented in Fig. 6a and 6b, respectively. The original signal is depicted in pink, and the output of the SG filter is shown in red. The figures demonstrate that the proposed SG filtering algorithm with an adaptive window length effectively smoothens the signal curve while preserving the original signal characteristics. The adaptive algorithm calculates the optimal window length for each position in the spectrum based on the signal features, in contrast to the original SG filtering algorithm, which employs a fixed window length. This approach enables filtering to preserve more original features while effectively decreasing noise and avoiding signal distortion. The results of this study suggest that the proposed algorithm is a promising tool for effectively filtering and analyzing complex beam spectra.

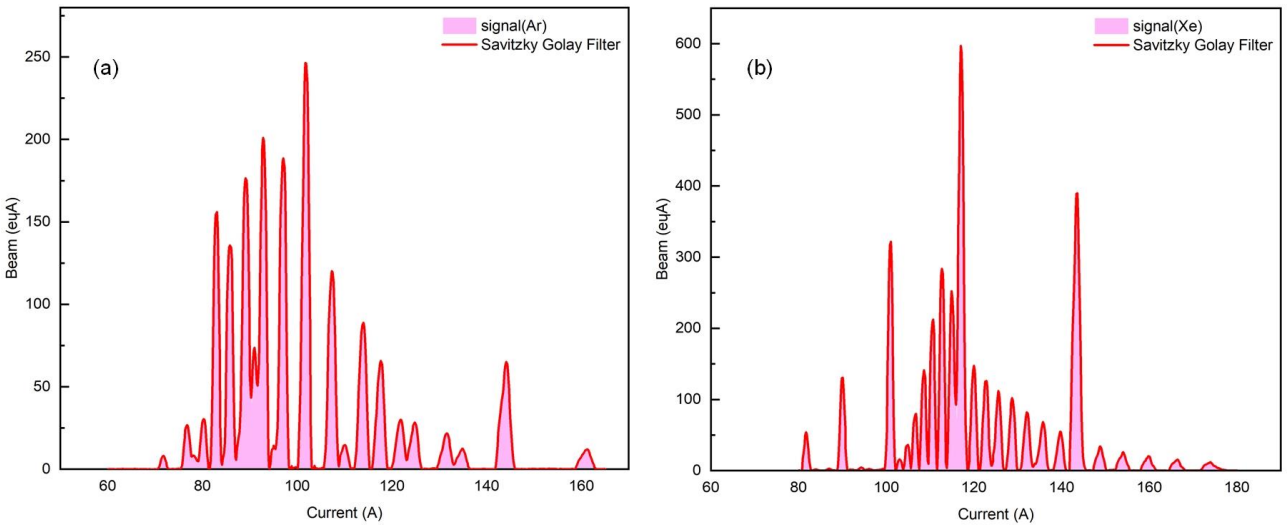


**Fig. 6** (a) Original Ar beam signal and SG filter with adaptive window length output. (b) Original Xe beam signal and SG filter with adaptive window length output.

In a previous study [42], the standard deviation of the first-order derivative was introduced as a metric to assess the effect of data smoothing. Specifically, a minor standard deviation of the first-order derivative indicated that the signal was smoothed to remove noise. To compare the effectiveness of the proposed SG filtering algorithm with adaptive window length against the original SG filtering algorithm, we examined the trend of the first-order derivatives in the original signal, the output of the ordinary SG filtering algorithm, and the output of the proposed algorithm. Fig. 7a and Fig. 7b depict the first-order derivative variations for the Ar and Xe beams, respectively. The blue line represents the first-order derivative trend of the original signal with respect to the number of data points. The red line represents the trend of the first-order derivative of the signal filtered using the original SG filtering algorithm, whereas the green line represents the trend using the proposed algorithm. To ensure a fair comparison, we used the same parameter values for the original SG filter as those initially used in the proposed algorithm, namely, $n=2$ and $N=5$. The results demonstrate that the proposed algorithm outperforms the original SG filtering algorithm by producing a smoother first-order derivative trend, indicating better noise reduction and signal preservation.

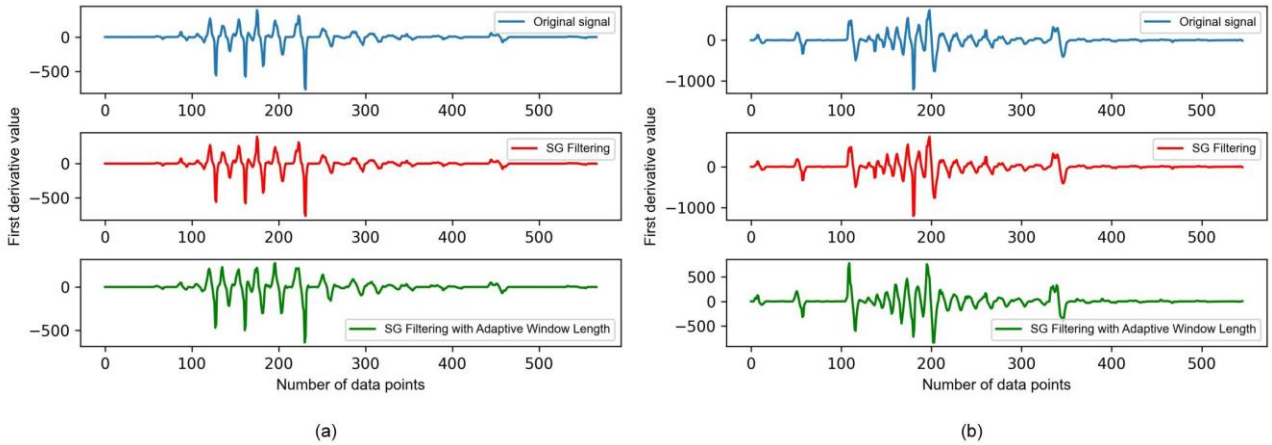


**Fig. 7** This study proposes an adaptive window length Savitzky-Golay (SG) filter algorithm and compares its performance with that of the original SG filter algorithm and the original signal. The trend of the first derivative of the output results was analyzed for both Ar and Xe beams. Specifically, we examine the variation in the first derivative with the number of data points for each algorithm. (a) shows the variation of the first derivative of the Ar beam with respect to the data points. (b) shows the same for the Xe beam.

The standard deviations of the first-order derivatives of the three signals were computed to compare the performance of the proposed SG filtering algorithm with the adaptive window length against that of the original SG filtering

algorithm and the original signal. Table 1 presents the results of the study. The table shows that The standard deviation of the first-order derivative decreases as the signal is filtered, indicating that noise is effectively reduced.

**Table 1** The standard deviation of the first derivative of the signal for the three cases.

| | The original signal | The ordinary SG filter algorithm | The SG filter algorithm with adaptive window length |
|---|---|---|---|
| Ar | 91.4372 A | 90.5672 A | 82.1846 A |
| Xe | 152.5697 A | 151.7765 A | 147.0195 A |

Specifically, the proposed algorithm achieved standard deviations of 82.1846 and 147.0195 A for the Ar and Xe beam spectra, respectively, whereas the original SG filtering algorithm yielded standard deviations of 90.5672 A and 151.7765 A for the same spectra. The proposed algorithm yields a minor standard deviation compared with the original SG filtering algorithm, indicating better signal preservation. These findings support the claim that the proposed algorithm offers a better denoising and smoothing effect on the beam spectra than the original methods evaluated in this study.

### 3.2 Performance analysis of improved AMPD algorithm

This study aims to evaluate the peak-seeking performance of the improved automatic multiscale peak detection (AMPD) algorithm by analyzing two beam spectra, as illustrated in Fig. 8. The Ar beam spectral peak detection is presented in Fig. 8a and Fig. 8b, whereas the Xe beam results are displayed in Fig. 8c and Fig. 8d. Owing to the denser charge-state distribution in the Xe beam spectrum, more peaks appear within a specific dipole magnet current variation range, leading to considerably greater detection difficulties than those encountered in the Ar beam spectrum. Nonetheless, the improved detection results of the AMPD algorithm agree with those obtained from physical experience, indicating its relative robustness to different beam spectra without the need for additional hyperparameters owing to the physical empirical threshold employed for peak filtering. Furthermore, the improved AMPD algorithm exhibited excellent performance in detecting the spectral peaks for high-charge states and dense distributions. The findings of this study suggest that the improved AMPD algorithm is a promising tool for accurate and efficient peak detection in complex spectra, particularly in scenarios involving highly charged states and dense distributions.

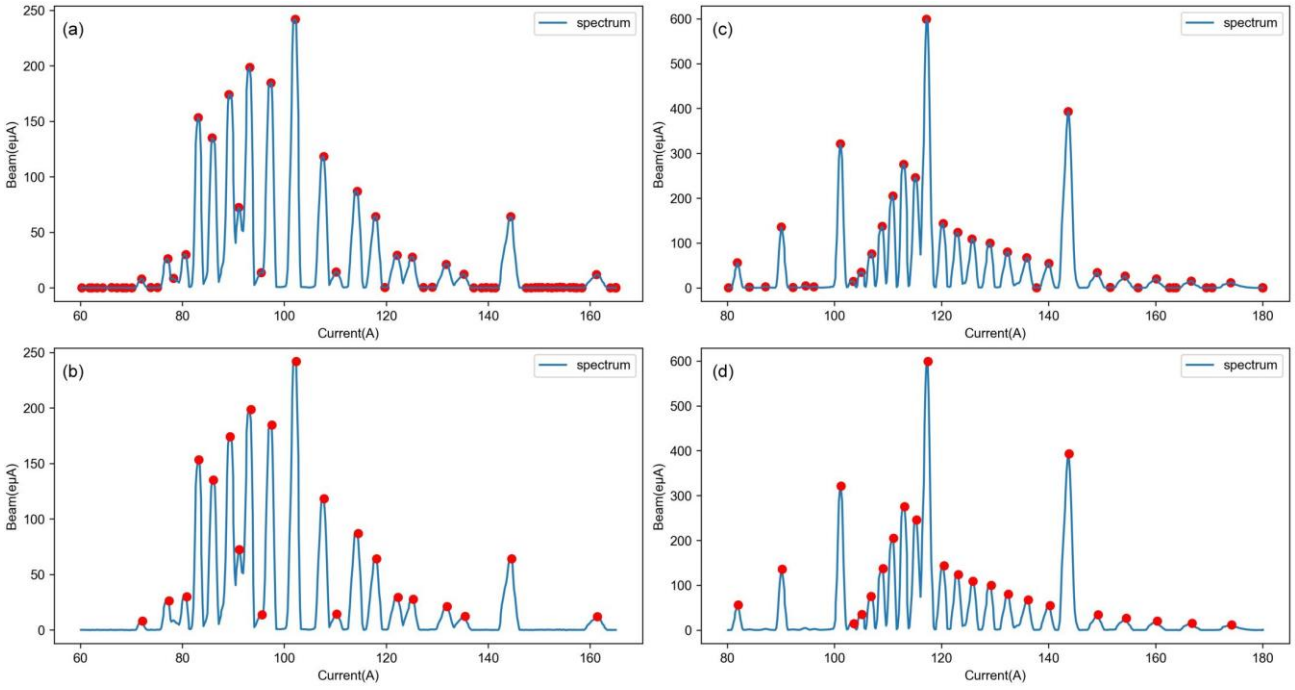


**Fig. 8** The improved AMPD algorithm proposed in this study was compared with the original AMPD algorithm

for peak detection results. (a) and (c) show the peak detection results of the original AMPD algorithm for the Ar and Xe beams, respectively. In contrast, (b) and (d) show the proposed improved AMPD algorithm results for the Ar and Xe beams, respectively.

### 3.3 Performance analysis of greedy matching algorithm based on relative distance

The original greedy algorithm yields a global solution that is merely a stack of all the local optimal solutions and may not be the optimal global solution. An IQR-based anomaly detection mechanism using distance encoding was proposed to address this issue. First, the distance-encoding output of the initial greedy algorithm based on the ion species should be classified. Next, we performed IQR-based anomaly detection and visually represented the anomalous values using box-line plots. The anomaly detection results for the Ar and Xe spectra are shown in Fig. 9. Specifically, Fig. 9a illustrates the anomaly detection results for the Ar beam spectra, revealing that only one anomaly appeared in the identification results of the main gas. In contrast, Fig. 9b shows the results for the Xe beam, where we identified two anomalous values for identifying the main gas.

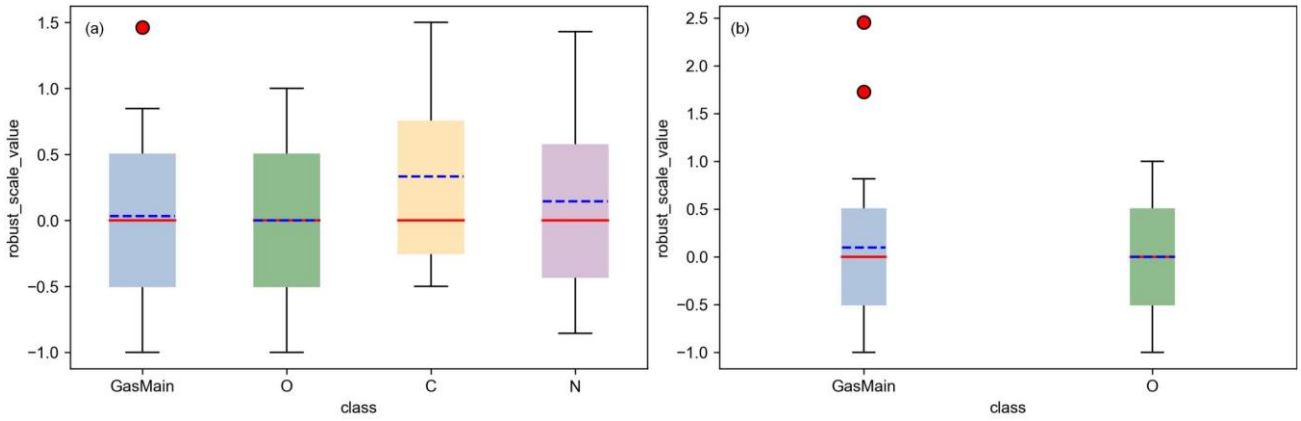


**Fig. 9** This figure illustrates the output results of the Ar and Xe beams obtained using the interquartile range (IQR) anomaly detection method. In the figure, the red dots represent the anomalous values, the red line segment denotes the median, and the blue dashed line indicates the arithmetic means for each category. Precisely, (a) displays the box plot of the Ar beam, and (b) illustrates the box plot of the Xe beam.

A secondary judgment was made after removing outliers to ensure the accuracy of the final identification results. Specifically, in the same spectrum with the same ion mass, the dipole magnet currents and charge states of the two adjacent peak positions must satisfy the empirical formula expressed in Eq. (12):

$$\frac{I_{q-1}}{I_q} = \sqrt{\frac{q}{q-1}} \tag{12}$$

The actual value of the $I_{q-1}/I_q$ was compared with the theoretical value of $\sqrt{q}/\sqrt{q-1}$. The accuracy of each matching term for each ion species was separately calculated as the difference between the actual and theoretical values. The root mean square error (RMSE) was subsequently computed and averaged for each ionic species. A lower RMSE implies a higher accuracy of the algorithm. The formula for calculating the RMSE is shown in Eq. (13) [43].

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{n}\left(Y_i - f\left(x_i\right)\right)^2} \tag{13}$$

The RMSE values of the matching results for the Ar and Xe beams are presented in Table 2, where $Y_i$ represents the actual value, and $f\left(x_i\right)$ represents the theoretical value in Eq. (13). To evaluate the results, we consider the algorithm acceptable when the RMSE < 0.1 A and excellent when the RMSE < 0.05 A [44, 45].

**Table 2** The root mean square error of the matching results of Ar and Xe beams

| Class | GasMain | C | N | O |
|---|---|---|---|---|
| RMSE（Ar） | 0.0678 A | 0.0034 A | 0.0028 A | 0.0024 A |
| RMSE（Xe） | 0.0880 A | —— | —— | 0.0115 A |

According to the data in Table 2, all the results fall within the acceptable range. The algorithm performed excellently in all categories except GasMain. This suggests that the proposed algorithm satisfies the accuracy requirements for beam spectrogram identification and has practical value in engineering applications.

### 3.4 Beam charge state spectrogram identification interface test

A software interface integrating data acquisition, peak finding, and spectrum recognition was developed. The interface, shown in Fig. 10, allows users to specify the start and end scan ranges and scan steps. Before peak seeking, users must specify the mass number of the main gas and the extracted high voltage of the ion source.

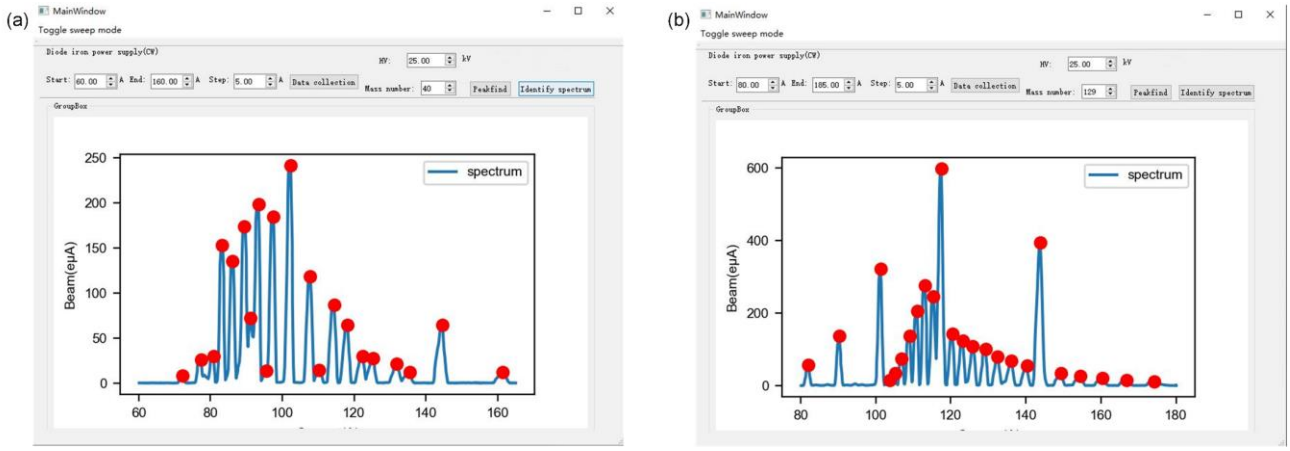


**Fig. 10** Using simple button clicks, an operator can easily set the beam parameters on the user interface and perform spectrum scanning, peak identification, and automatic spectrum recognition functions. (a) illustrates the primary interface for Ar beam spectrum identification, and (b) shows the primary interface for Xe beam spectrum identification.

Clicking the "Identify spectrum" button in the main interface will open the identify spectrum interface, which displays the kind of ion and charge-state corresponding to each peak, as shown in Fig. 11. Users can zoom in on the interface to view the details. The identification results obtained using the system proposed in this study meet the accuracy requirements for identifying the Ar (Fig. 11a) and Xe (Fig. 11b) beams.

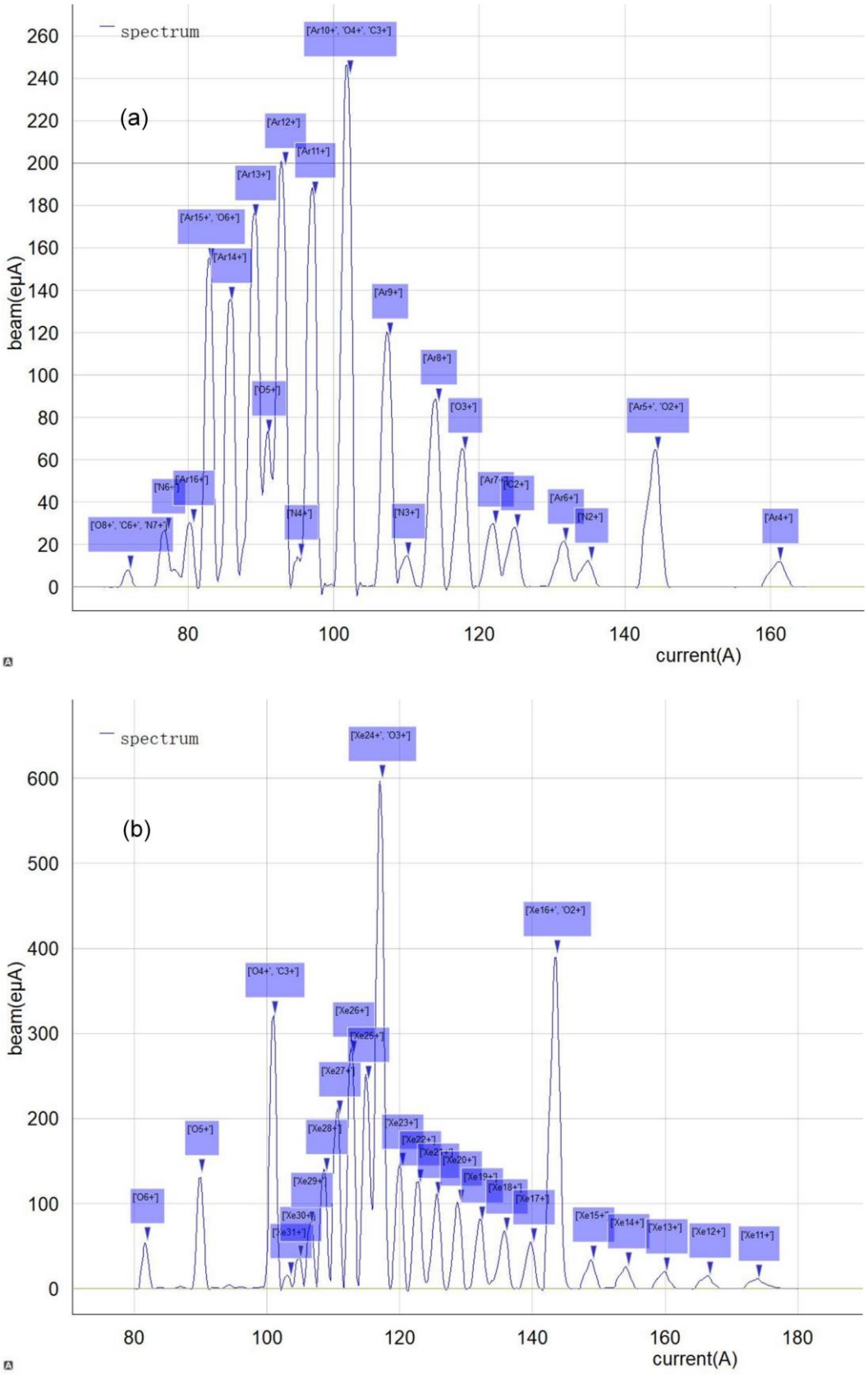


**Fig. 11** This figure depicts the performance evaluation of the beam charge-state distribution spectrogram identification system for the Ar and Xe beams. (a) Spectrum recognition results for the Ar beam and (b) spectrum recognition results for the Xe beam.

## 4 Conclusion

In this study, a novel spectral recognition system for beam charge-state distribution of high-charge-state Electron Cyclotron Resonance (ECR) ion sources was designed and developed. It enables automatic spectrum recognition using three algorithms: Savitzky-Golay (SG) filtering with adaptive window length, improved automatic multiscale peak detection (AMPD), and greedy matching based on relative distance.

The system achieved optimized smoothing with the special SG filtering algorithm introduced in this study and accurate peak detection with the improved AMPD algorithm. The proposed greedy matching algorithm effectively identified spectral peaks and their ion charge states. The system offers a user-friendly software interface and a real-time display of the results. This system effectively and accurately identified the beam spectra of Xe and the spectra of the mass numbers below Xe.

The advantages of the developed automatic spectrum recognition system are as follows:

1. The SG filtering algorithm with adaptive window length improved smoothing performance.
2. The improved AMPD algorithm eliminates the need for pre-setting hyperparameters and effectively rejects false peaks.
3. The greedy matching algorithm with relative distance, augmented by the IQR anomaly-detection mechanism, ensures accurate and efficient matching, thereby overcoming the limitations of the greedy algorithm.
4. The user-friendly software interface enables easy parameter specification and real-time display of spectral identification results.

This automatic spectrum recognition system improves the commissioning efficiency of the ECR ion source and achieves a precise charge-state ion beam injection, making it a valuable contribution to high-charge-state ECR ion source research.

## Author Contributions:

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Rui Wang and Cheng Qian. Yu-Hui Guo contributed to the key revisions of the manuscript. Peng Zhang and Jin-Dou Ma guided some physical aspects of the experiments. The first draft of the manuscript was written by Rui Wang and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

## Data Availability Statement:

The data that support the findings of this study are openly available in Science Data Bank at https://www.doi.org/10.57760/sciencedb.08240 and https://cstr.cn/31253.11.sciencedb.08240.

## References


1. R. Geller, ECRIS: The electron cyclotron resonance ion sources. ANNU REV NEUROSCI. 40,1 15-44 (1990). doi:10.1201/9780203758663
2. Q. Wu, Y.G. Liu, J.L. Liu et al., Design of a high current ion source for an electromagnetic isotope separator. Nucl. Tech. **46**(03), 030202. (2023). doi: 10.11889/j.0253-3219.2023.hjs.46.030202
3. D. Leitner, C.M. Lyneis, T. Loew et al., Status report of the 28 GHz superconducting electron cyclotron resonance ion source. VENUS. REV SCI INSTRUM. **77**(3), 03A302. (2006). doi: 10.1063/1.2149298
4. T. Nakagawa, Y. Higurashi, J. Ohnishi et al., First results from the new RIKEN superconducting electron cyclotron resonance ion source. REV SCI INSTRUM. **81**(2), 02A320. (2010). doi: 10.1063/1.3259232
5. H. Zhao, L. T. Sun, J. W. Guo et al., Intense highly charged ion beam production and operation with a superconducting electron cyclotron resonance ion source. PHYS REV ACCEL BEAMS. **20**(9), 094801.

(2017). doi: 10.1103/physrevaccelbeams.20.094801
6. H.F. Xiao, Q.X. Zhang, H.Y. Tan et al., The study of a neutron spectrum unfolding method based on Particle Swarm Optimization combined with Maximum Likelihood Expectation Maximization. Nucl. Sci. Tech. **34**(4), 60. (2023). doi: 10.1007/s41365-023-01200-8
7. Z.Y. Yao, Y.S. Xiao, J.Z. Zhao. Dose reconstruction with Compton camera during proton therapy via subset-driven origin ensemble and double evolutionary algorithm. Nucl. Sci. Tech. **34**(4), 59. (2023). doi：10.1007/s41365-023-01207-1
8. R.Y. Wu, C.R. Geng, F. Tian et al., GPU-accelerated three-dimensional reconstruction method of the Compton camera and its application in radionuclide imaging. Nucl. Sci. Tech. **34**(4), 52. (2023). doi：10.1007/s41365-023-01199-y
9. S.C. Zheng, Q.Q. Pan, H.W. Lv et al., Semi-empirical and semi-quantitative lightweight shielding design algorithm. Nucl. Sci. Tech. **34**, 43. (2023). doi: 10.1007/s41365-023-01187-2
10. H. Li, J. Shi, L. Li et al., Novel wavelet threshold denoising method to highlight the first break of noisy microseismic recordings. IEEE T GEOSCI REMOTE. **60**, 1-10 (2022). doi: 10.1109/tgrs.2022.3142089
11. L. Hu, L. Wang, Y. Chen et al., Bearing Fault Diagnosis Using Piecewise Aggregate Approximation and Complete Ensemble Empirical Mode Decomposition with Adaptive Noise. SENSORS. **22**(17), 6599. (2022). doi: 10.3390/s22176599
12. K. Hasan, S.T. Meraj, M.M. Othman et al., Savitzky–Golay Filter-Based PLL: Modeling and Performance Validation. IEEE T INSTRUM MEAS. **71,** 1-6. (2022). doi: 10.1109/tim.2022.3196946
13. V.C.B. Sousa, C. Scalo. A Legendre Spectral Viscosity (LSV) Method Applied to Shock Capturing for High-Order Flux Reconstruction Schemes. J COMPUT PHYS. **460**, 111157. (2022). doi: 10.1016/j.jcp.2022.111157
14. K. Chen, A. Badji, S. Laghrouche et al., Polymer electrolyte membrane fuel cells degradation prediction using multi-kernel relevance vector regression and whale optimization algorithm. APPL ENERG. **318**, 119099. (2022). doi: 10.1016/j.apenergy.2022.119099
15. F. Wu, X. Kong, C. Xu, Test on Stochastic Block Model: Local Smoothing and Extreme Value Theory. J SYST SCI COMPLEX. **35**(4), 1535-1556. (2022). doi: 10.1007/s11424-021-0154-9
16. G. Wang, Y. Wang, Y. Min et al., Blind Source Separation of Transformer Acoustic Signal Based on Sparse Component Analysis. ENERGIES. **15**(16), 6017. (2022). doi: 10.3390/en15166017
17. H. Kennedy, Recursive digital filters with tunable lag and lead characteristics for proportional-differential control. IEEE T CONTR SYST T. **23**(6), 2369-2374. (2015). doi: 10.1109/tcst.2015.2399436
18. E.N. Nishida, O.O. Dutra, L.H.C. Ferreira et al., Application of Savitzky-Golay digital differentiator for QRS complex detection in an electrocardiographic monitoring system. in *2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. (Rochester, MN, USA, 2017). doi: 10.1109/memea.2017.7985881
19. D. Suescún-Díaz, H.F. Bonilla-Londoño, J.H. Figueroa-Jimenez, Savitzky–Golay filter for reactivity calculation. J NUCL SCI TECHNOL-T. **53**(7) 944-950. (2016). doi: 10.1080/00223131.2015.1082949
20. K. Pandia, S. Ravindran, R. Cole et al., Motion artifact cancellation to obtain heart sounds from a single chest-worn accelerometer. in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. (Dallas, TX, USA,2010). doi: 10.1109/icassp.2010.5495553
21. J.M. Gregoire, D. Dale, R.B. Van Dover, A wavelet transform algorithm for peak detection and application to powder x-ray diffraction data. REV SCI INSTRUM. **82**(1), 015105. (2011). doi: 10.1063/1.3505103
22. J.S. Park, S.W. Lee, U. Park, R peak detection method using wavelet transform and modified shannon energy envelope. J HEALTHC ENG. (2017). doi: 10.1155/2017/4901017

23. N. Mtetwa, L.S. Smith, Smoothing and thresholding in neuronal spike detection. NEUROCOMPUTING. **69**(10-12), 1366-1370. (2006). doi: 10.1016/j.neucom.2005.12.108
24. P. Sedighian, A.W. Subudhi, F. Scalzo et al., Pediatric heart sound segmentation using Hidden Markov Model. in *2014 36th annual international conference of the ieee engineering in medicine and biology society*. (Chicago, IL, USA, 2014). doi: 10.1109/embc.2014.6944869
25. C.L. Forgy, Rete: A fast algorithm for the many pattern/many object pattern match problem, Readings in Artificial Intelligence and Databases. 547-559. (1989). doi: 10.1016/0004-3702(82)90020-0
26. A. Bouchet, Greedy algorithm and symmetric matroids. MATH PROGRAM. **38**, 147-159. (1987). doi: 10.1007/bf02604639
27. Z. He, S. Deng, X. Xu et al., A fast greedy algorithm for outlier mining. in Advances in Knowledge Discovery and Data Mining: 10th Pacific-Asia Conference, PAKDD 2006, (Singapore, April 9-12, 2006). doi: 10.1007/11731139_67
28. Z. Zhao, M. Zhou, S. Liu, Iterated greedy algorithms for flow-shop scheduling problems: A tutorial. IEEE T AUTOM SCI ENG. (2021). doi: 10.1109/tase.2021.3062994C.T. Ryan, R.L. Smith, A greedy algorithm for finding maximum spanning trees in infinite graphs. OPER RES LETT. **50**(6), 655-659. (2022). doi: 10.1016/j.orl.2022.10.004
29. C.T. Ryan, R.L. Smith, A greedy algorithm for finding maximum spanning trees in infinite graphs. OPER RES LETT. 50(6), 655-659. (2022). doi: 10.1016/j.orl.2022.10.004
30. X. Han, Y. Han, Q. Chen et al., Distributed flow shop scheduling with sequence-dependent setup times using an improved iterated greedy algorithm. Complex Adapt. Syst. Model. **1**(3), 198-217. (2021). doi: 10.23919/csms.2021.0018
31. H.P. Vinutha, B. Poornima, B.M. Sagar, Detection of outliers using interquartile range technique from intrusion dataset. in *Information and Decision Sciences: Proceedings of the 6th International Conference on FICTA*. (Bhubaneswar, Odisha, 2018). doi: 10.1007/978-981-10-7563-6_53
32. A. Hoshiar, T.A. Le, F. Amin et al., A novel magnetic actuation scheme to disaggregate nanoparticles and enhance passage across the blood–brain barrier. NANOMATERIALS-BASEL. **8**(1), 3. (2017). doi: 10.3390/nano8010003
33. M. Tokarska, M. Frydrysiak, J. Zięba, Electrical properties of flat textile material as inhomegeneous and anisotropic structure. J MATER SCI-MATER EL. **24**, 5061-5068. (2013). doi: 10.1007/s10854-013-1524-4
34. S. Sepúlveda, P. Reyes, A. Weinstein. Visualizing physiological signals in real-time. in *Proc. of the 14th Python in Science Conf.* (Austin, Texas, 2015). doi: 10.25080/majora-7b98e3ed-01c
35. S.W. Bai, X.F. Yang, S.J. Wang et al., Commissioning of a high-resolution collinear laser spectroscopy apparatus with a laser ablation ion source. Nucl. Sci. Tech. **33**(1), 9. (2022). doi: 10.1007/s41365-022-00992-5
36. M. Sadeghi, F. Behnia, R. Amiri, Window selection of the Savitzky–Golay filters for signal recovery from noisy measurements. IEEE T INSTRUM MEAS. **69**(8), 5418-5427. (2020). doi: 10.1109/tim.2020.2966310
37. B.H. Prasetio, E.R. Widasari, H. Tamura. Automatic multiscale-based peak detection on short time energy and spectral centroid feature extraction for conversational speech segmentation. in *6th International Conference on Sustainable Information Engineering and Technology 2021*. (Malang, 2021) doi: 10.1145/3479645.3479675
38. H. Zhao, B. Tian, B. Chen, Robust stable iterated unscented Kalman filter based on maximum correntropy criterion. AUTOMATICA. **142**, 110410. (2022). doi: 10.1016/j.automatica.2022.110410
39. C. Leys, C. Ley, O. Klein et al., Detecting outliers: Do not use standard deviation around the mean, use

absolute deviation around the median. J EXP SOC PSYCHOL. **49**(4), 764-766. (2013). doi: 10.1016/j.jesp.2013.03.013

40. X. Wan, W. Wang, J. Liu et al., Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC MED RES METHODOL. **14**, 1-13. (2014). doi: 10.1186/1471-2288-14-135
41. Y. Luo, W.P. Lin, P.P. Ren et al., A simulation study of a windowless gas stripping room in an E//B neutral particle analyzer. Nucl. Sci. Tech. **32**(7), 69. (2021). doi: 10.1007/s41365-021-00909-8
42. G. Taubin, Curve and surface smoothing without shrinkage. in *Proceedings of IEEE international conference on computer vision*. (Cambridge, MA, USA ,1995). doi: 10.1109/iccv.1995.466848
43. S. Yuan, Review of root-mean-square error calculation methods for large deployable mesh reflectors. INT J AEROSPACE ENG. (2022). doi: 10.1155/2022/5352146
44. V. Savalei, The relationship between root mean square error of approximation and model misspecification in confirmatory factor analysis models. EDUC PSYCHOL MEAS. **72**(6), 910-932. (2012). doi: 10.1177/0013164412452564
45. N.V. De Giesen, S.C. Steele-Dunne, J. Jansen et al., Double-ended calibration of fiber-optic Raman spectra distributed temperature sensing data. SENSORS. **12**(5), 5471-5485. (2012). doi: 10.3390/s120505471